Fast Adaptive Tracking Based on Fusion Particle Filter Algorithm

Jubao Qu[1, a], Hongtao Liang[2, b]

[1]School of mathematics and computer, Wuyi University Wuyishan, Fujian, China, Qjbok@qq.com

[2]School of mathematics and computer, Wuyi University, Wuyishan, Fujian, China

[a]Qjbok@qq.com, [b]Lianght@qq.com

Abstract: To lock and track the target quickly, a novel fast target tracking algorithm is proposed, which combines adaptive scale Harris corner detection target, SIFT adaptive matching target and particle filter algorithm. This algorithm can extract highly discriminant features of the target adaptively, and make the feature pair rotate and scale zoom. The brightness change can keep invariance, and maintain a certain degree of stability for affine transformation, angle change and noise. It can also recognize targets well in the case of confusion and occlusion. The simulation results show that the algorithm can still accurately identify and track targets under multiple external disturbances, and has certain application value.

1. Introduction

With the rapid development of network technology, people pay more and more attention to remote monitoring, guidance, social security and modern transportation by using cameras. How to quickly lock and track a given target between multiple cameras has become a hotspot in the field of computer vision application. Document [1,2] proposes a continuous tracking method in the case of overlapping horizons between cameras, but overlapping horizons limit the effective monitoring range; Document [3,4] uses the motion of the quasi-target object. Trajectory realizes consistent discriminant tracking of objects, but the accuracy of this method depends on the length of the trajectory. Document [5,6] regards the time and space of the target entering and leaving the field of vision as constraints, and combines the color-space histogram matching method to achieve continuous tracking of objects, but it is greatly affected by the speed, position and illumination of the target motion. Document [7,8] Target handover between multiple cameras is achieved by using the distance between the target center and the camera visual field boundary of synchronous video. The disadvantage of this algorithm is that it is difficult to obtain the visual field boundary between two cameras when the overlap area is small, or even impossible to obtain it. Moreover, the algorithm only uses a straight line to distinguish targets, when there are several cameras at the same time; Literature [9] removes the "outdated" feature information caused by noise by real-time updating of SIFT features of targets, which improves the stability of features and the accuracy of matching. However, this method can only be used in fixed scenes with a single camera. For this reason, this paper proposes a fast-tracking algorithm which combines adaptive scale Harris corner detection, SIFT adaptive matching and particle filter algorithm. The simulation results show that the algorithm can still track targets in the presence of external interference. High efficiency, not easy to lose and so on.

2. Establishment of Moving Object Model

Usually the camera is fixed, and the motion of moving object is random. To track the same object in multiple cameras is the problem of the target's transition from one camera field of view to another camera field of view. Because the scene illumination of each camera is different, the background is different, and the camera parameters are not changed. Similarly, it is very difficult to track. For this reason, this paper uses the median filter method which has strong noise ability and fast detection

speed to build the background mean image, obtains the target area by background difference, and then removes the background and shadow to get the pure target, because the background will confuse the characteristics of the target, and the shadow will change the shape of the target.

2.1 Establishment of background mean image

Using median filtering method, the continuous L-frame image in the video is stored in the image buffer, and then the average value of the pixels in the same position of the L-frame image is calculated to obtain the background mean image of the k-frame:

$$\Psi_k(x, y) = median\{\Phi_k(x, y),...,\Phi_{k-L}(x, y)\} \tag{1}$$

In formula: $\Phi_k(x, y)$ is the k-frame image with X and Y coordinates of pixels.

The background image should be updated in real time because of the change of background pixels caused by moving objects and changing illumination. When the difference between the pixels corresponding to the current frame and the background mean image is less than a threshold, the background pixels will remain unchanged, otherwise the following dynamic update formula will be adopted:

$$\Psi_{k+1}(x, y) = \begin{cases} \Psi_k(x, y), & if \quad |\Phi_{k+1}(x, y) - \Psi_k(x, y)| < \delta \\ \alpha \cdot \Phi_{k+1}(x, y) + (1-\alpha) \cdot \Psi_k(x, y), otherwise \end{cases} \tag{2}$$

In the formula: δ is the frame difference threshold, $\alpha \in (0,1)$ is the update coefficient, the larger α, the faster background update speed.

2.2 Motion target acquisition

The background frame difference method is used to subtract the current frame image from the background image. If the difference is less than the threshold aaaa, it is considered as the background image and is juxtaposed by 0. On the contrary, it is the target image. The target image model is as follows.

$$\Gamma_{k+1}(x, y) = \begin{cases} 0, & if \quad |\Phi_{k+1}(x, y) - \Psi_k(x, y)| < \delta \\ \Phi_{k+1}(x, y), & otherwise \end{cases} \tag{3}$$

2.3 Achieving Pure Object

In target detection, shadow often accompanies target at the same time, which causes great obstacles to the subsequent extraction of target feature information. In this paper, HSV color feature space which can reflect gray and color information more accurately is used for shadow clipping. The shadows formed by the scales cause different changes in the HSV component. The shadows darken the brightness, and the V component will decrease greatly; the saturation S component and the chroma H component will change little. The background difference between the k-frame video image aaaaaa and the background image sssss is made on the H, S and V components. If the difference is less than a threshold, the shadows will be placed side by side with 0. Otherwise, the detection model is as follows.

$$\Gamma_k(x, y) = \begin{cases} 0, & if \ \alpha^V \le \Phi I_k^V(x, y) / \Psi_k^V(x, y) \le \beta^V \\ & and \ |\Phi_k^S(x, y) - \Psi_k^S(x, y)| < \delta^S \\ & and \ |\Phi_k^H(x, y) - \Psi_k^H(x, y)| < \delta^H \\ \Gamma_k(x, y), & otherwise \end{cases} \tag{4}$$

In formula: I^H, I^S, I^V and B^H, B^S, B^V represent the H, S and V components of the pixel values of the current frame and background frame at (i, j), $0 \le \alpha^V \le \beta^V \le 1$, δ^H and δ^S are the

thresholds of the chroma and saturation components respectively, α^V and β^V are the thresholds of the brightness components. In experiment: $\alpha^V = 0.1, \beta^V = 0.8, \delta^H = 0.2, \delta^S = 0.4$.

3. Adaptive Scale Harris Corner Detection Target

The classical SIFT algorithm can ensure that the extreme points detected at each scale are the key points in each scale, but it cannot guarantee that these key points can appear at different scales. When the target is too small, the number of key points detected is too small to meet the three or more requirements (sub-pixel accuracy) required for precise positioning. An adaptive scale Harris corner detection method is used to compensate for the insufficient number of matching points in SIFT algorithm. When the number of matching points is not more than 3 pairs, Harris corner detection method is started to find other key points. Harris corner detection method has the advantages of simple calculation, stable operator, uniform and reasonable corner features, and can quantitatively extract different sizes. The algorithm is as follows.

Let $\Phi_x(x, y)$ and $\Phi_y(x, y)$ be the first-order partial derivatives of image I (x, y) in X and Y directions respectively, then the corner metric and M matrix at different scales are:

$$\xi_x^2(x, y, \sigma_k) = G(x, y, \sigma_k) * \Phi_x^2(x, y) \tag{5}$$

$$\xi_y^2(x, y, \sigma_k) = G(x, y, \sigma_k) * I_y^2(x, y) \tag{6}$$

$$\xi_{xy}(x, y, \sigma_k) = G(x, y, \sigma_k) * I_x(x, y) I_y(x, y) \tag{7}$$

$$\mathbf{M}(x, y, \sigma_k) = \begin{bmatrix} L_x^2(x, y, \sigma_k) & L_{xy}(x, y, \sigma_k) \\ L_{xy}(x, y, \sigma_k) & L_y^2(x, y, \sigma_k) \end{bmatrix} \tag{8}$$

Among them, $G(x, y, \sigma_k)$ is a Gauss function and $\mathbf{M}(x, y, \sigma_k)$ is an M matrix at scale σ_k (calculated by adaptive scale formula 8).

Let $\det(\mathbf{M}(x, y, \sigma_k))$ be the determinant value of the matrix, λ the empirical value (0.05 in the experiment), $\text{trace}(\mathbf{M}(x, y, \sigma_k))$ the trace of the matrix, and Hariss corner response function at scale σ_k is:

$$R(x, y, \sigma_k) = \det(\mathbf{M}(x, y, \sigma_k)) - \lambda \cdot \text{trace}^2(\mathbf{M}(x, y, \sigma_k)) \tag{9}$$

When the Harris corner response value $R(x, y, \sigma_k)$ of a point D (x, y) is greater than the threshold μ (0.01 in the experiment), it can be determined that the point is the corner point. To ensure that Harris corner is not the key point detected by SIFT, the new detected corner points are selected or rejected: when the corner coordinates are the same as the key point coordinates, the corner points are lost, otherwise the key points of SIFT are retained; Keep corners and join the queue as key points.

4. Continuous target tracking

To extract highly discriminant features from the target, which can keep invariance to rotation, scale scaling, brightness change, affine transformation, perspective change, noise, and robust target recognition in chaotic and occlusive situations, SIFT algorithm has the above advantages. It establishes two-dimensional scale space and DoG (Difference of Gaussian) Gauss difference space, detects the location, scale and direction features of local key points, and matches the SIFT features of two images using SIFT feature vectors to achieve tracking.

4.1 Generating image scale space

Convoluting two-dimensional image, I (x, y) with Gauss kernel $G(x, y, \sigma)$ to generate scale space L.

$$\xi(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \tag{10}$$

In the formula: $G(x, y, \sigma) = \dfrac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$ is the Gaussian kernel function; (x, y) is the coordinate of the pixel; σ is called the scale space factor.

4.2 Constructing DOG Gauss difference scale space

Convoluting different scales of Gauss difference kernels with image I (x, y).

$$\begin{aligned} \Pi(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma))I(x, y) \\ &= \xi(x, y, k\sigma) - \xi(x, y, \sigma) \end{aligned} \tag{11}$$

4.3 Detection and Accuracy of Key Points

Comparing each point of each image in DOG scale space with its 8 neighborhood points of the same scale and 18 neighborhood points of the upper and lower adjacent scales as the candidate points of key points. Then, by fitting the three-dimensional quadratic function, the position and scale of key points are accurately determined, and the edge unstable points and scales are removed. Low contrast point to enhance matching stability and enhance anti-noise ability.

4.4 Determining the main direction of the key point

The gradient of each pixel in the neighborhood window (16 *16 pixels) centered on the key point is weighted by the Gauss function, and the gradient direction in the window is counted by the histogram. The direction corresponding to the maximum value in the histogram is the main direction of the key point. The gradient magnitude ϕ and direction φ of the key point are as follows.

$$\phi(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \tag{12}$$

$$\varphi(x, y) = \tan^{-1} \frac{(L(x, y+1) - L(x, y-1))}{(L(x+1, y) - L(x-1, y))} \tag{13}$$

Among them, the scale used by L is the scale where the key points are located.

4.5 Generate SIFT eigenvectors

To ensure the invariance of rotation, the coordinate axis is rotated clockwise from θ^o to the main direction of the key point. The 8*8 rectangular window centered on the key point is evenly divided into four 4*4 sub-rectangles. Eight directional gradient histograms are counted for each sub-rectangle, and the length of the eigenvector is normalized. A SIFT eigenvector of 4*8=32 dimensions is generated. At this time, SIFT feature vectors have removed the effects of scale, rotation and illumination changes.

4.6 SIFT feature matching

When the SIFT feature vectors of the two images are generated, the Euclidean distance of the feature vectors of the key points is used as the similarity measure of the key points in the two images. A key point is extracted from the previous image and the first two key points closest to the Euclidean distance in the latter image are found. If the ratio is less than a threshold τ, the point pair closest to

Euclidean distance is the matching point pair. Otherwise, the matching point pair is not considered to be the matching point pair.

5. Fast Tracking Using Particle Filter

Particle filter arithmetic originates from Monte Carlo idea, approximates the target with the help of particle set, and then checks the probability density function. This arithmetic can be used in any mode of state model. Its essence is to operate the integral calculation in Bayesian estimation with the help of Monte Carlo idea, assuming the system equation and the probability density function. The equation of state is.

$$x_k = f(x_{k-1}) + w_{k-1}, \quad z_k = h(x_k) + v_{k-1} \tag{14}$$

Among them, Wk is system noise and Vk is process noise. Both are independent and identically distributed noise sequences. The posterior probability density is.

$$q(x_k \mid x_{k-1}^i, \ z_k) = p(x_k \mid x_{k-1}^i) \tag{15}$$

In the standard particle filter, to solve the problem conveniently, the prior probability distribution is chosen by default as the optimal recommended distribution function. The standard particle filter algorithm is described as follows:

5.1 Initialization

k = 0 and P (x0) were used to extract N sample points:
$\{x_0^i, \ i = 1, 2, \cdots N\}$

5.2 Importance weight calculation

Importance sampling (k=1,2, N), according to the following formula random sampling of a group of particles, the number of particles is N.

$$\omega_k^i = w_{k-1}^i \cdot \frac{p(x_k^i \mid x_{k-1}^i) p(z_k \mid x_{k-1}^i)}{q(x_k^i \mid x_{0:k-1}^i, \ z_{1:k-1})} \tag{16}$$

If the one-step posterior state distribution is used, the above formula can be simplified to $\omega_k^i = \omega_{k-1}^i \cdot p(z_k \mid x_k^i)$ and the normalized importance weight is $\overline{\omega}_k^i = \omega_k^i / \sum_{j=1}^{N} \omega_k^j$.

5.3 Resampling

According to the size of each normalized weight $\overline{\omega}_k^i$, the samples $\overline{x}_{0:k}^i$ were duplicated or discarded, and N samples $p(x_{0:k}^i \mid z_{1:k})$ approximating $x_{0:k}^i$ distribution were obtained. $\omega_k^i = \overline{\omega}_k^i = 1/N, \quad i = 1, \cdots, N$.

5.4 Output result

The particle set $\{x_{0:k}^i : i = 1, 2, \cdots N\}$ output by the algorithm can approximate the posterior probability and the expectation of function $f_k(x_{0:k})$.

$$\overline{p}(x_{0:k} \mid z_{1:k}) = \frac{1}{N} \sum_{i=1}^{N} \delta_{x_{0:k}^i}(dx_{0:k}) \tag{17}$$

$$E(f_k(x_{0:k})) = \frac{1}{N} \sum_{j=1}^{N} f_k(x_{0:k}^j) \qquad (18)$$

5.5 Judge whether the tracking is overs

If it is, quit the algorithm, if not, return to step B.

6. Experimental simulation and evaluation comparison

The experimental system chooses outdoor surveillance scene without overlapping field of view of multiple cameras for simulation experiment. The video resolution is 480 x 340, the frame rate of image acquisition is 25f/s. The experimental software environment is Win2010, VC++, MATLAB and OpenCV. It runs on a PC with Core2(3.99) GHz and 4G memory.
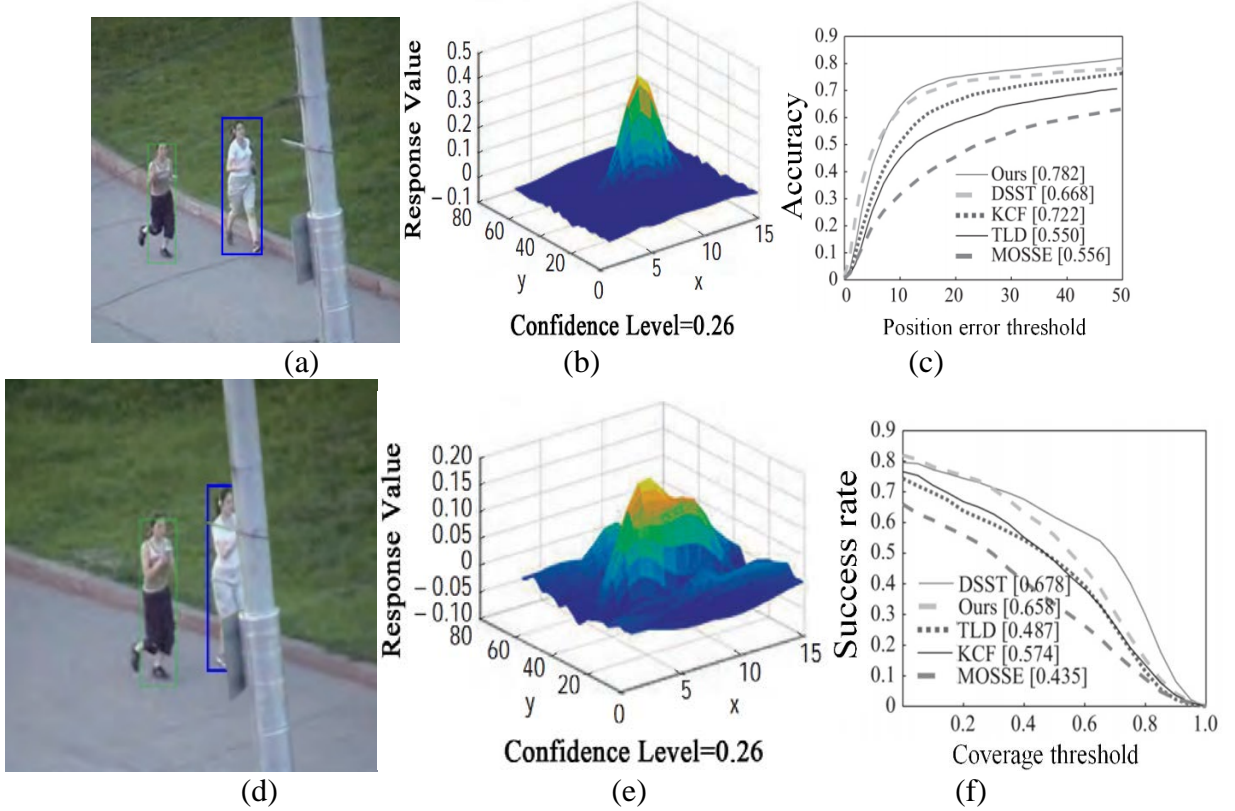

Fig. 1 Correct Rate Contrast Graph of Scale Change Video Sets

Fig. 1 (a, d) is different from two frames in the Jogging 1 video sequence in the OTB-50 standard data set. Fig. 1 (b, e) is their corresponding confidence changes. Fig. 1 (c, f) is the correct rate comparison when scale and shape change. From this we can see that, in the shape and scale changes, the algorithm in this paper has not only the pyramid model to cope with the scale changes, but also the update and correction mechanism of the model, which is no inferior to DSST.

Tab. 1 Comparison of Tracking Time-consuming and Accuracy

	Check matching points	Correct matching points (%)	Time consuming (s)	Accuracy rate
Document [2] algorithm	925	682	12	81.2
Document [3] algorithm	1625	1352	31	85.0
Our algorithm	91	84	2.9	93.6

Using three algorithms to track this video, the time-consuming and accuracy are shown in Table 1. It can be seen from the table that the algorithm in this paper extracts the key points from pure targets and adopts the adaptive SIFT algorithm, so the overall calculation is less, the matching speed is fast and the accuracy is high. Although the illumination intensity and the perspective of the two videos are quite different, it often still exists. There is mutual occlusion between targets, but each tracking index is superior to other algorithms and achieves better results.

7. conclusion

In this paper, after studying the SIFT feature matching algorithm, an improved SIFT feature matching algorithm with adaptive function is proposed. The target is detected by the scale-adaptive Harris corner, and the fusion algorithm of fast-tracking target by particle filter algorithm is used to improve the usual algorithm in target scale change, occlusion, deformation and so on. Robustness in the case of multi-target successive tracking is achieved successfully. The algorithm has been tested on several data sets, and the simulation results show that the tracking effect is good, and there is a good space for further research.

Acknowledgments

References

[1] Lowe David G. "Distinctive image features from scale-invari-ant keypoints," International Journal of Computer Vision,2004, pp.91-110

[2] Bernardini C, Silverston T, Festor O.MPC: popularity-based caching strategy for content centric networks. //IEEE International Conference on Communications, 2013, pp.3619-3623.

[3] KE Tao, ZHANG Yongjun. The application of SIFT features in the automatic matching for low altitude aerial images. Science of Surveying and Mapping, 2009, pp.23-26

[4] Lowe D G. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision,2004, pp.91-110

[5] BRADSKIG R. Computer vision face tracking for use in a perceptualuser interface//Proceedings of Fourth IEEE Workshop Applications of Computer Vision. Berlin: IEEE, 1998, pp.214-219

[6] Chen Guanxiong, Han Xiaopu, Wang Binghong.Multi-level scaling properties of instant-message communica-tions. Physics Procedia, 2010, pp.1897-1905.

[7] Suga A, Fukuda K,Takiguchi T, Ariki Y. Object recognition and segmentation using SIFT and graph cuts//In:Proc. of the 19th International Conference on Pattern Recognition.Tampa, USA:IEEE, 2008, pp.254-259.

[8] LIN Hai-Feng,MA Yu-Feng,SONG Tao. Research on Object Tracking Algorithm Based on SIFT, ACTA AUTOMATICA SINICA August,2010, pp.1204-1208

[9] NIU Enguo, CHENG Yongmei, PAN Quan, Fast Algorithm for Automatic Generating Field of View Line of Multiple Cameras. Application Research of Computers,2007, pp.297-299